DOI: 10.13382/j.jemi.B2407724

基于知识蒸馏的空间通道双自编码器无监督异常检测*

梁 宵 陈 莹

(江南大学轻工过程先进控制教育部重点实验室 无锡 214122)

摘 要:在工业检测场景下,按照是否引入正常样本中不存在的异常,可以将异常检测问题分为结构异常检测和逻辑异常检测两类,逻辑异常检测对网络的全局理解能力提出了更高的要求。针对现有无监督异常检测模型在结构异常上已有较好的检测精度,但无法适应逻辑异常检测需求的问题,提出一种包含空间聚合模块和通道聚合模块的双自编码器结构,主要由 3 部分组成。首先设计了并行空空间通道双自编码器架构,从空间和通道两个方向得到包含全局信息的特征向量,提升网络的长程依赖关系;其次设计一个选择性融合模块,融合双自编码器信息,放大包含重要信息的特征,以进一步提高对逻辑异常的表达能力;最后提出在自编码器与学生网络的损失函数中加入余弦损失,避免网络对单个像素差异过于敏感,从而关注于全局差异。在 MVTec LOCO AD 数据集上进行实验,逻辑异常检测精度达到 89.4%,结构异常检测精度达到 94.9%,平均检测精度 92.1%,超越了基线方法和其他无监督缺陷检测方法,验证了方法的有效性和优越性。

关键词: 无监督异常检测;逻辑异常检测;并行双 AE;选择性融合模块;融合余弦相似性损失

中图分类号: TP391.4; TN911.7 **文献标识码**: A **国家标准学科分类代码**: 520.2

Knowledge distillation based spatial channel dual autoencoders for unsupervised anomaly detection

Liang Xiao Chen Ying

(Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China)

Abstract: In industrial detection scenario, according to whether anomalies that do not exist in normal samples are introduced, anomaly detection problems can be divided into two categories: structural anomaly detection and logical anomaly detection. Logical anomaly detection places higher demands on the global understanding ability of the network. Faced with the problem that the existing unsupervised anomaly detection model has a good detection accuracy on structural anomalies, but cannot meet the requirements of logical anomaly detection, a dual autoencoder structure consisting of spatial reunion module and channel reunion module is proposed. Our method consists of three components: Initially, the parallel space channel dual autoencoder architecture is introduced, by obtaining feature vectors containing global information from spatial and channel directions, the long-range dependencies of the network is improved. Secondly a selective fusion module is designed to fuse the information of the dual autoencoder and amplify features containing important information to further improve the ability to express logical anomalies. Lastly cosine loss is proposed to the loss function between autoencoder and student network to avoid the network being sensitive to individual pixel differences, so as to focus on global differences. We conducted experiments on MVTec LOCO AD dataset, and achieved 89.4% in logical anomaly detection accuracy, 94.9% in structural anomaly detection accuracy, and 92.1% in average detection accuracy, surpassing the baseline method and other unsupervised defect detection methods, verifying the effectiveness and superiority of the method.

Keywords: unsupervised anomaly detection; logical anomaly detection; parallel dual AE; selective fusion module; cosine similarity fused loss

收稿日期:2024-07-29 Received Date: 2024-07-29

*基金项目:国家自然科学基金(62173160)项目资助

0　引　言

随着成像技术、深度学习等技术的发展,异常检测在医学图像[1]、视频监控和工业质量检测[2-3]等各种任务中有了重要的应用。在工业异常检测的背景下,异常样本通常是稀缺并且外观难以预测的,获取异常样本的高昂成本使得有监督的方法难以应用于真实的工业场景。同时,相比于异常样本,由于正常图像具有更易获取、数量丰富、获取成本低等优势,基于正常样本建模的无监督异常检测方法,得到了越来越多的关注。

在工业异常检测领域,可能出现的异常缺陷种类多种多样,例如在生产一个装有特定数量和种类螺丝的螺丝袋时,可能出现的缺陷就包括螺丝上的划痕和裂纹、袋中螺丝数量或种类的错误等。为了适应多变的实际生产环境和生产要求,将工业生产过程中的异常缺陷,根据其是否引入正常数据中不存在的异常,分为结构异常和逻辑异常两类。局部异常,即指出现在局部区域的、正常区域中不存在的异常,例如出现在螺丝上的划痕和弯曲,类似的还包括织物产品上的破损和油污、坚果上的裂纹等等,注重于检测单一的产品。而对于没有出现异常的特征,而是由于正常特征的错误排列组合,造成不满足生产要求的异常缺陷称为逻辑异常,例如袋中螺丝数量或种类的错误,这类由正常物品的错误排序或丢失重要成分造成的异常,类似的还有电路接线错误、药瓶液面偏移或缺少等。逻辑异常在生产过程中也会不可避免地出现,本文注重于研究同时适用于结构异常和逻辑异常的检测方法。

由于结构异常检测的应用范围广泛,之前的无监督异常检测方法已经提出了不少研究方法和研究项目,并取得了不错的效果。早期工作主要是基于单分类的方法[4-5],即训练模型,将数据映射到特征空间的一个最小球面,使网络提取正常数据中的共同特征,测试点到球中心点的距离作为异常得分。目前主流的方法可以分为 3 类:基于重建的方法、基于特征表示的方法和基于知识蒸馏的方法。基于重建的方法[6-7]通常训练自编码器(autoencoder, AE)、生成对抗网络等生成模型对图像进行重建。由于生成模型只被训练为生成正常图像,测试时异常样本重建前后的差异会比正常图像大,由此判断异常。基于特征表示的方法[8-10]首先使用在大型数据集上预训练的网络提取特征,然后对正常特征进行建模来辨别异常特征。最近的研究表明,基于知识蒸馏的无监督模型[11-14]通过使用在大型数据集上预训练的教师网络,已经能在结构异常缺陷检测的任务上实现不错的效果。

经过近几年的发展,结构异常缺陷的检测技术已经有了长足的进步,Bergmann 等[15]首先提出了更具挑战性的逻辑异常缺陷检测问题,不少研究人员也转而关注于解决这类问题。研究表明,相比于结构异常,逻辑异常的检测更具挑战性,这是因为划痕、裂纹及纹理等结构异常不需要模型理解对象的高级语义特征,而仅使用局部的知识即可进行判断,属于低级异常。而逻辑异常检测需要模型发现图像是否违反样本潜在的逻辑约束,因此逻辑异常的检测对模型的全局理解能力提出了更高的要求。

之前的方法缺乏足够大的感受野,仅基于局部语义判断正常。这就导致之前针对于结构异常缺陷检测的方法,在遇到逻辑异常缺陷时往往表现不佳,例如在需要检测的对象包含固定数量的长螺丝、短螺丝和螺帽这类具有全局上下文限制的复杂场景中,模型无法在关注每个产品质量的同时保证产品的数量、位置不违反逻辑要求,这类正常物体出现在错误的位置或物体丢失的缺陷,很可能会被误判为正常。如何提高对逻辑异常的检测能力,也成为了研究的热点,逻辑异常检测的方法也可以分为基于图像重构、基于特征提取和基于知识蒸馏的 3 类。

在基于图像重构的方法中,研究人员在针对结构异常的方法上进行改进,通过提高或限制部分网络的感受野,提高模型对逻辑异常的检测能力。Guo 等[16]提出了一种模板引导的分层特征还原方法,通过一个正常模板库来引导异常特征恢复为正常特征。Dai 等[17]使用去除自注意力机制的去噪扩散概率模型,限制其感受野,以生成保留局部结构、忽略全局结构的特定异常图像。

在基于特征提取的方法中,Kim 等[18]通过对少量标注图片,将图像分割成多个部件,并储存进 3 个记忆库当中,匹配分割出的部件的个数、特征和图片每个块的细粒度特征。Chen 等[19]观察到缺陷检测中常常把少量的困难正样本错误地识别为异常样本,提出构建一个基于原型的稳健决策边界,并在查询集和模板集两个方向上互相探索异常,从而能够地捕捉逻辑异常。

基于知识蒸馏的方法中,Bergmann[15]首次提出了全局上下文异常检测(global context anomaly detection, GCAD)方法,设计了全局-局部两个模型,并引入了自动编码器。自动编码器需要对图像压缩得到潜在空间,利用其用有全局感受野的优势对逻辑异常的语义信息进行建模。Yao 等[20]设计了一个语义瓶颈层,以增强局部-全局特征的对应性。Zhang 等[21]在反向蒸馏(reverse distillation, RD)[22]的网络基础上设计了本地和全局的双学生网络模型,通过引入全局上下文压缩模块和上下文亲和损失,来进一步增强全局学生捕捉长距离依赖关系的能力。Batzner 等[23]提出了一个高效的异常检测网络 EfficientAD,通过限制学生教师网络的感受野,在全局分支中扩大全局特征与局部特征的差异,进一步提高了

逻辑异常检测的性能。

尽管 EfficientAD 针对逻辑异常检测进行了改进,实验结果表明该模型在逻辑异常样本上的检测结果依然比起结构异常有较大差距。造成这个问题的原因可能是全局分支中,自编码器网络没有得到足够丰富的全局上下文信息,或者没有得到充分的利用。并且本文注意到,在训练过程中需要缩小学生网络和自编码器网络的输出差异,来辅助自编码器生成更清晰的特征,直接使用欧式距离计算每个特征的距离,可能导致自编码器关注于局部的特征差异。早期的学生网络对自编码器指导意义并不大,将两者进行像素级别的对齐甚至可能造成学生网络的性能下降,对结构异常和逻辑异常的检测效果都有损害。

针对上述问题,为了实现更好地检测逻辑异常,本文设计了空间聚合模块(spatial reunion module, SRM)和通道聚合模块(channel reunion module, CRM),分别得到包含全局空间和通道理解的特征,并且提出在 EfficientAD 的全局分支中采用并行双自动编码器的结构,来丰富模型对局部特征的上下文建模。并通过设计一个选择性融合模块(selective fusion module, SFM),灵活地融合通道聚合 AE 和空间聚合 AE 的输出,进一步增强特征表示,提高网络对逻辑异常的检测效果。

同时又为了保证逻辑异常检测的前提下,不损害网络结构异常检测效果,本文提出在计算学生网络和自编码器网络之间损失函数时考虑方向差异,引入融合余弦相似性(cosine similarity fused, CSF)的损失函数,使网络不会对每个像素的差异过于敏感,更好地捕捉重要的上下文信息。在 MVTec LOCO AD 数据集上进行的实验验证了本研究方法的有效性,大量提升了逻辑异常的检测精度,并且小幅优化了结构异常检测精度,实现了逻辑异常和结构异常的兼顾。

1 本文方法

1.1 总体框架

本文方法的总体框架如图 1 所示。图 1 中使用不同的底色区分了局部和全局两个分支,并用黑色实线表示全局分支网络的前进方向,蓝色实线表示局部分支网络的前进方向,测试时前进方向与训练时一致,不同点在于用黑色和蓝色的虚线表示了异常图的生成过程。并以逻辑异常样本为例,展示了本文方法对于逻辑异常缺陷检测的效果。整体网络可以分为局部和全局两个分支,分别针对结构异常缺陷和逻辑异常缺陷。其中局部分支由学生教师两个网络组成,学生-教师网络采用块描述网络(patch description network, PDN)[23],每一个输出的特征向量只描述周围像素的区域,这样的好处是限制了学生和教师网络的感受野,保证图像上某一部分的异常不会引起较远部分的响应,适合捕捉局部异常。教师网络使用在 ImageNet 上预训练的 WideResNet-101 网络蒸馏得到,使其具有丰富的表达的能力,能够表征正常和异常的特征,而学生网络采用随机初始化的参数,只在正常样本上训练,只能表征正常的特征。

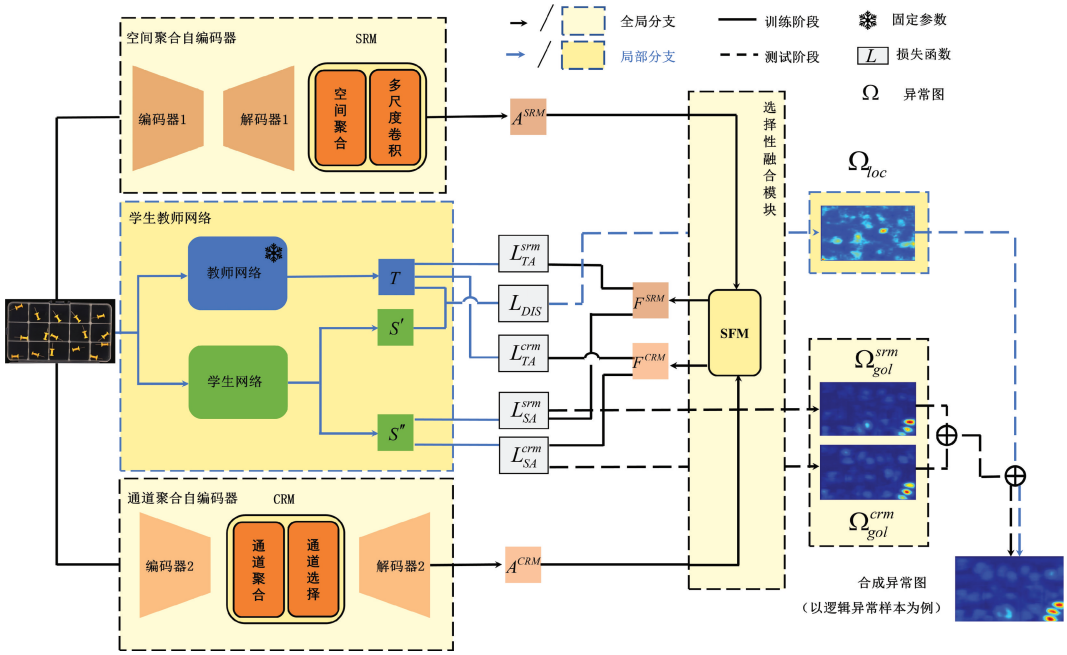


图 1 方法总体框架

Fig. 1 Overall method framework

针对逻辑异常检测,全局分支由并行双自编码器网络和选择性融合模块组成。为了充分利用自编码器的感受野的优势得到丰富、有利的长程语义信息,本文在两个自编码器网络中分别设计了 SRM 和 CRM 模块,从空间和通道两个方向得到互补的全局信息,从两个方向提升网络的长程依赖关系。空间聚合 AE 的输出 A^{SRM} 和通道聚合 AE 的输出 A^{CRM}。再通过 SFM 模块进行选择性地融合,进一步增强特征表示。

学生网络的训练通过对齐和教师网络的输出之间的欧氏距离实现。由于逻辑异常没有引入正常数据中不存在的特征,且只在正常样本上训练的 PDN 学生网络只有局部的感受野,因此测试时,其对正常区域和逻辑异常区域的表征和教师应该没有明显差异;而在结构异常上二者输出差异较大,从而使学生-教师网络专注于检测结构异常,得到结构异常图 Ω_{loc}。

自编码器网络的训练由教师-自编码器(teacher-autoencoder, T-A)和学生-自编码器(student-autoencoder, S-A)两部分组成。在 T-A 部分,分别对齐两个自编码器和教师网络的输出。这里和蒸馏网络的训练相似,不同的是,由于自编码器的输出比较模糊,难以重建细粒度的表征[24],直接使用教师网络和自编码器的输出差异作为异常图会造成假阳性检测。而学生网络输出通常也无法实现教师网络的精细表示,使用学生网络训练自编码器能够减轻假阳性检测的问题。因此在 S-A 部分中,将学生网络的输出通道数翻倍,原始的一半通道 S' 仍用于学生-教师网络,翻倍的一半通道 S'' 用于和自编码器对齐。并且通过 SFM 模块,选择性融合两个自编码器的输出,得到更有益的特征 F^{SRM}、F^{CRM}。其中 AE 和教师网络对齐时使用欧氏距离,自编码器和学生网络 S'' 对齐时,使用本文提出的 CSF 损失函数。这样做的好处是,一方面避免了学生网络的性能下降,一方面提高学生网络和自编码器的训练过程的稳定性,实验表明这样的设计对模型的逻辑异常和结构异常检测效果都有提升效果。测试时,由于学生网络和自编码器网络感受野的差异,通过计算学生网络和两个自编码器网络的差异得到逻辑异常图,分别得到逻辑异常图 Ω_{gol}^{srm} 和 Ω_{gol}^{crm}。将结构异常图和逻辑异常图进行归一化后合并,得到的组合异常图能够实现结构异常和逻辑异常检测的兼顾。

1.2　空间聚合模块

将自编码器最后输出的特征作为输入 I 输入 SRM 模块,通过多尺度卷积和聚合操作两部分,生成具有长距离空间信息的新特征,SRM 如图 2 所示。

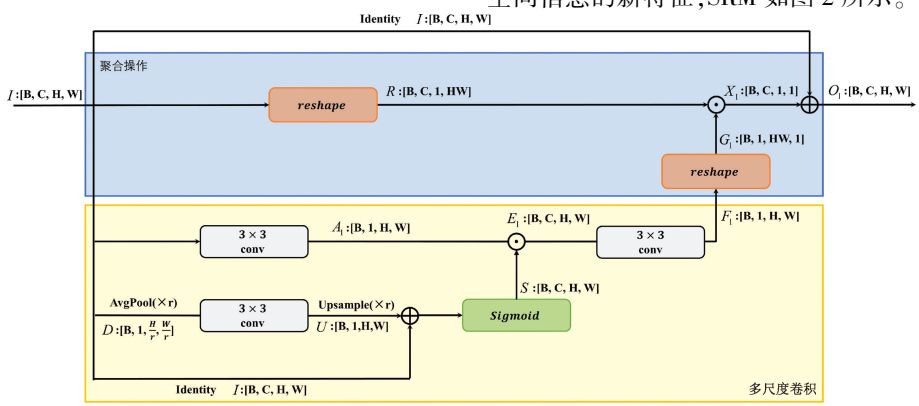


图 2　空间聚合模块

Fig. 2　Spatial reunion module

首先,为了得到每个空间位置的语义信息,本文在两个不同的尺度空间中进行卷积操作,即原始的尺度空间和经过下采样后的尺度空间。通过平均池化操作,得到具有较大的感受野的特征空间:

$$D = AvgPool_{\times r}(I) \tag{1}$$

通过一个 3×3 的卷积层和双线性插值操作,将其恢复到原始特征的尺寸,使其能与原始的输入特征相融合,以保留原始信息:

$$U = UpSample_{\times r}(Conv_{3\times 3}(D)) \tag{2}$$

将得到的 U 经过线性激活模块,与原始尺度空间的特征 I 相乘,作为参考来始特征空间中的特征转换过程:

$$A_1 = Conv_{3\times 3}(I) \tag{3}$$

$$F_1 = Conv_{3\times 3}(A_1 \cdot \sigma(U + I)) \tag{4}$$

其中,σ 为 sigmoid 函数,公式为:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

得到的特征图中,每个空间位置不仅可以自适应地考虑其周围的信息背景,将其作为潜在空间的嵌入,在原始尺度空间的响应中发挥标量的作用,有效扩大了卷积层的视场。

此时的感受野考虑了每个空间位置周围多尺度的上下文信息,但是得到的感受野仍然有限,只能在局部空间

起作用,要捕捉长距离的依赖关系依然比较困难。因此,将多尺度卷积部分的输出 F_1 重塑为[B, 1, $H \times W$, 1];在聚合操作部分,将原始的输入特征 I 重塑为[B, C, 1, $H \times W$],并将二者相乘,得到空间视场的全局理解 $X_1 \in R^{B\times C\times 1\times 1}$:

$$X_1 = reshape(F_1) \cdot reshape(I) \tag{6}$$

将得到的 U 经过线性激活模块,与原始尺度空间的特征 I 相乘,作为参考来指导原始特征空间中的特征转换过程。

最后,将得到的特征与原始的输入特征通过残差连接融合,使网络能更轻松地训练:

$$O_1 = I + X_1 \tag{7}$$

模块消融实验可提供可视化结果,从中可以看出空间聚合模块能够有效地得到空间层面的长程语义信息,这不仅得益于其在不同尺寸空间上的卷积操作进行多尺度信息编码带来的丰富感受野变化,也得益于通过矩阵乘法操作,得到位置较远的像素的相互关系,从而实现空间层面的全局上下文理解。

1.3 通道聚合模块

众所周知,神经网络输出的特征中,每一个通道都可以看作对输入数据的某种特定语义的响应,而不同的通道响应之间并不是完全独立的,存在着相互依赖关系。在 CRM 中,本文对通道间的依赖关系进行建模,通过整合所有特征图的相关通道,获得通道维度的长程语义信息,从而有选择性地强调重要的通道,抑制无用的冗余特征的通道,具体实现如图 3 所示。

以自编码器生成的中间层特征作为输入 I,首先使用两个 1×1 的卷积层,将通道数压缩为 1,获得了包含全局通道的特征,与输入的特征相乘,用于得到包含全局通道理解的向量:

$$A_2 = Conv_{1\times 1}(Conv_{1\times 1}(I)) \tag{8}$$

$$E_2 = A_2 \cdot I \tag{9}$$

之后两个全连接层,用于建模通道之间的依赖关系,学习各个通道的重要性,自适应调整各个通道的权重,这个权重即反映了该通道在当前任务中的重要性,实现对含有重要信息的通道放大:

$$X_2 = \sigma(FC_2(FC_1(E_2))) \tag{10}$$

通道聚合模块通过矩阵乘法和通道的挤压、变化操作,模拟通道间的依赖关系,实现了通道层面的全局理解。

1.4 选择性融合模块

从两个自编码器模块得到的特征能够从空间、通道两个角度获取样本的长程依赖关系,有效提高性能。然而,如果直接将两个模块的输出相加进行特征融合,可能会造成潜在的信息丢失,为此本文设计了选择性融合模块,如图 4 所示,对于相加后的特征 $M \in R^{H\times W\times C}$,使用全局平均池化层和全连接层,进行挤压和混合。

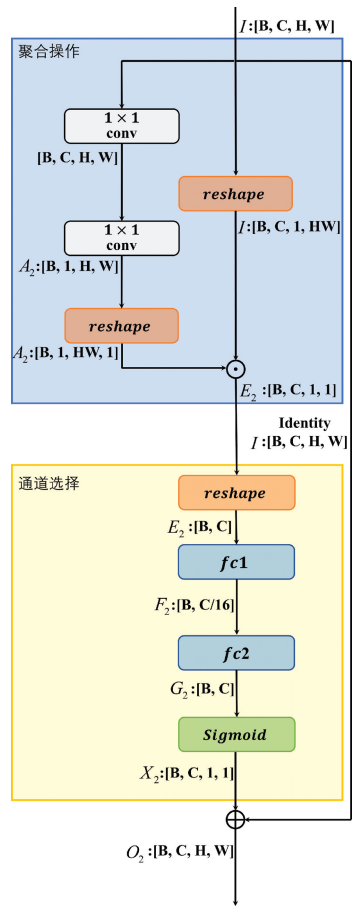


图 3 通道聚合模块

Fig. 3 Channel reunion module

经过 1×1 的卷积层和 softmax 层后,得到的特征 $N_1 \in R^{1\times 1\times C}$ 和 $N_2 \in R^{1\times 1\times C}$ 可以表示每个通道的重要性,通过通道乘法与输入特征 O_1、O_2 分别相乘进行调整,得到输出特征 Y_1 和 Y_2。

1.5 损失函数

损失由两部分组成,即蒸馏损失和自编码器损失。

1)蒸馏损失

蒸馏损失用于学生-教师网络中,使用欧式距离判断教师网络 T 和学生网络的前半通道的输出的 S' 是否对齐。

$$L_{ST} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}(T_{i,j}(I) - S'_{i,j}(I))^2 \tag{11}$$

式中:I 为输入的正常样本;$T_{i,j}(I)$ 和 $S'_{i,j}(I)$ 为教师网络的输出和学生网络前半通道的输出;H、W 分别为 $T_{i,j}(I)$、$S'_{i,j}(I)$ 的高度和宽度。

为了避免学生网络对于异常区域的过度泛化,在欧氏距离的基础上使用困难损失和惩罚损失。困难损失

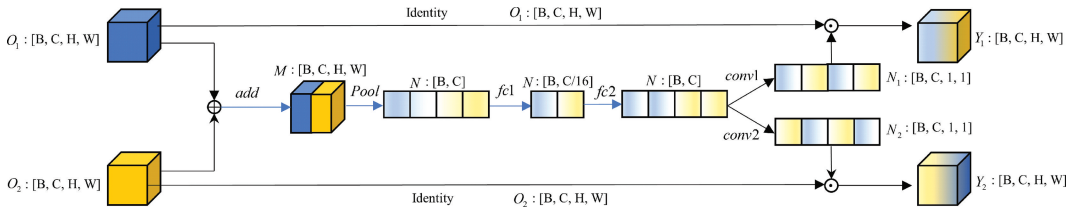


图 4　选择性融合模块

Fig. 4　Selective fusion module

L_{HARD} 在 L_{ST} 的基础上只选择最远的距离作为损失进行反向传播，设定困难损失分位数 $p_{hard}=0.999$。即困难损失 L_{HARD} 是从 L_{ST} 中，只取前 0.1%大的值参与反向传播。

另外，通过计算学生网络对预训练数据集中的样本的输出作为惩罚项，得到惩罚损失，希望进一步抑制学生对分布外数据的泛化能力。

$$L_{PEN}=\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}(S_{i,j}(P))^2 \tag{12}$$

式中：P 为从 ImageNet 中随机采样图片。

最终学生网络的蒸馏损失由困难损失和惩罚损失组成[23]：

$$L_{DIS}=L_{HARD}+L_{PEN} \tag{13}$$

2）自编码器损失

自编码器损失由 T-A 和 S-A 两部分组成。在 T-A 部分，同样使用欧式距离判断教师网络 T 和自编码器网络 A 的输出是否对齐。

$$L_{TA}=\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}(T_{i,j}(I)-A_{i,j}(I))^2 \tag{14}$$

式中：$A_{i,j}(I)$ 为自编码器的输出。由于自编码器重建的特征往往无法恢复精细纹理的表示[25]，直接使用教师和自编码器网络之间的差异作为异常图，会造成假阳性误检。考虑到学生网络的输出会是比较模糊的正常特征，因此将学生网络的输出通道数翻倍，并将翻倍的一半通道用于使自编码器和学生网络对齐。

$$L_{SA}^{dis}=\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}(A_{i,j}(I)-S''_{i,j}(I))^2 \tag{15}$$

式中：$S''_{i,j}(I)$ 为学生网络对输入正常样本 I 的翻倍通道的输出。

然而，欧氏距离很难表达复杂特征空间的相似性，对于自编码器而言，更希望网络不关注于局部特征的差异，而能有全局上下文的长程理解。因此本文提出在计算欧氏距离的基础上，融合余弦相似性的损失函数，通过增加余弦损失，从角度和距离两个方面平衡两个网络之间的相似性，抑制个别像素的噪声干扰，并提高网络的鲁棒性。欧氏距离侧重于计算每个维度的像素差异，而余弦距离测量数据之间的几何角度，对于每个像素的数值变化不敏感，能更好地捕捉输出特征方向上的差异，余弦距离计算公式为：

$$L_{SA}^{\cos}=\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(1-\frac{A_{i,j}(I)\cdot S''_{i,j}(I)}{\|A_{i,j}(I)\|\ \|S''_{i,j}(I)\|}\right) \tag{16}$$

得到自编码器与学生网络之间的损失函数：

$$L_{SA}=L_{SA}^{dis}+\lambda L_{SA}^{\cos} \tag{17}$$

式中：λ 为可调的超参数。在并行双自编码器中，分别计算每个自编码器的损失，自编码器网络的总损失为：

$$L_{AE}=(L_{TA}^{srm}+L_{SA}^{srm})+(L_{TA}^{crm}+L_{SA}^{crm}) \tag{18}$$

总的训练损失为：

$$L_{TOTAL}=L_{DIS}+L_{AE} \tag{19}$$

测试时，将局部分支和全局分支的异常图进行归一化后结合得到最终的异常图。

2　实验结果及分析

2.1　实验设置

为了评价本文方法在结构异常和逻辑异常缺陷上的效果，本文在 MVTec LOCO AD[15] 数据集上进行实验。MVTec LOCO AD 数据集由 MVTec 公司于 2022 年推出，由工业检测场景中的 5 个对象类别，共 3 600 余张图片组成，每个类别都有一定的逻辑约束。如对于早餐盒类别的图片，不光要求盒中的食物不应有缺陷，即满足结构正常要求，同时要求每个样本中恰好有两个橘子和一个油桃，并且它们总是位于盒子的左侧。对于每一个类别，训练集和验证集都为正常图像，而测试集包含结构异常图像、逻辑异常图像和正常图像。

评价指标选择接收者操作特征曲线下的面积（area under the receiver operating characteristic curve，AUROC）来评估图像级精度。

本文将输入图像统一调整为 256×256 pixels。训练周期为 70 000 次，采用 Adam[26] 优化器，设置学习率为 1×10^{-4}，权重衰减为 1×10^{-5}，批次大小为 1，余弦损失的超参数 λ 设 0.5。

2.2　实验结果

将本文方法与现有的优秀的结构异常检测方法和针对逻辑异常检测的方法进行比较，比较的方法有

SPADE[8]、GCAD[15]、RD[22]、DSKD[21]、SINBAD[27]、STPM[14]、DADF[28]和基线方法 EfficientAD[23]。

实验结果如表1所示,本文方法相比于 EfficientAD 方法在逻辑异常/结构异常上精度分别提高了2.7%/0.8%,达到了89.4%/94.9%,平均精度提高了1.7%,达到了92.1%,超越了其他现有的方法,证明了本文方法的优越性。

表1　MVTec LOCO AD 数据集上的检测结果,每个类别的结果按照逻辑异常精度/结构异常精度展现,平均精度展示了逻辑异常平均精度/结构异常平均精度(L/S)以及全部平均精度

Table 1　The detection results on the MVTec LOCO AD dataset, the results for each category are reported as Logical anomalies/Structural anomalies, the overall averages are reported as Logical anomalies/Structural anomalies(L/S) and the average of all　　(%)

方法	早餐盒	果汁瓶	图钉	螺丝包	连接器	平均(L/S)	平均
SPADE	81.8/74.7	91.9/84.9	60.5/58.1	46.8/59.8	73.8/57.1	69.0/78.9	74.0
STPM	68.9/68.4	82.9/74.7	59.5/90.3	55.5/87.0	65.4/96.8	66.4/88.3	77.4
GCAD	87.0/80.9	**100**/98.9	97.5/74.9	56.0/70.5	89.7/78.3	86.0/80.7	83.4
SINBAD	**96.5**/87.5	96.6/93.1	83.4/74.2	**78.6**/92.2	89.3/76.7	88.9/84.7	86.8
RD	66.7/60.3	93.6/95.2	63.6/84.8	54.1/89.2	75.3/95.9	70.7/85.1	77.9
DADF	75.8/74.8	98.7/98.4	76.7/85.3	66.2/88.4	78.6/94.2	79.2/88.2	83.7
DSKD	—	—	—	—	—	81.2/86.9	84.0
EfficientAD	85.0/87.3	97.5/**99.8**	96.7/93.8	57.8/91.1	96.7/**98.3**	86.7/94.1	90.4
本文	85.5/**88.6**	99.6/**99.8**	**98.7/95.6**	65.9/**92.4**	**97.4**/98.2	**89.4/94.9**	**92.1**

图5和6所示分别为本文方法和基线方法、STPM、RD方法在逻辑异常样本和结构异常样本上的可视化结果对比,由于STPM和RD模型中没有设置局部分支与全局分支,对比时直接展示了它们最终输出的热力图。由于逻辑异常缺陷不易观察,直接展示真值图片对判断结果并不直观。因此,在本文的可视化对比实验中,对于逻辑异常图片,展示正常模板样本,并在异常样本中用方框标出了可以被视为逻辑异常缺陷的区域,以便更直观地观察,而对于结构异常,直接展示了对应的真值图片。从图5可以看出,在检测图钉盒中多出的图钉时,efficientAD的全局分支热力图没有定位到异常的格子,并导致合成的组合异常图错误,而STPM和RD方法虽然定位到了异常的区域,却在其他很多正常区域也有较高的响应。在检测早餐盒类中种类错误的水果时,efficentAD的全局分支图标注出了两个水果的位置,实际上3个位置的水果都有可能是缺陷位置,本文方法检测区域更加完整,考虑更全面,而STPM和RD方法都错误地关注于右侧坚果区域。在检测连接器类时,图例样本线缆颜色与接头的个数不匹配,efficientAD方法在全局分支和合成热力图中只标注出了线缆的一部分,STPM方法的标注不连续,且也有错误的响应,RD方法只标注了两边的结构,没有考虑到连接线错误的可能性。在螺丝包类中,本应包含一长一短两根螺栓的样本中,错误地包含了两根短螺栓,因此两根螺栓都应判断为异常,efficientAD、STPM和RD方法没有发现这个异常,本文方法将两根螺栓的位置都标注了出来。在检测液面过高的果汁瓶类时,efficientAD和RD虽然标注了液面,但也有错误的响应,并且分割不如本文方法完整,而STPM方法错误地标注了标签。从以上对比图可以看出,相比于STPM、RD方法和基线方法efficientAD,本文方法能更准确地识别出违反正常样本逻辑约束的异常,检测的区域更准确、完整。

图6为基线方法、STPM方法与本文方法在结构异常缺陷上的检测结果,在检测图钉盒中多出的异物时,efficientAD和本文方法的局部分支热力图都成功定位到了这个结构异常,但是全局分支热力图中,efficientAD错误的标注导致了合成的异常图多出了两处误检,而本文方法在全局分支中定位依然正确,所以在输出时减少了误检,RD的定位不如本文方法精确,STPM仍然有不少多余的响应。在检测早餐盒中出现的药片时,本文方法在局部分支和全局分支中检测效果都较efficientAD要好,最终组合异常图避免了对水果表皮的误检,相比于STPM也更加精准。在检测连接器中断裂的连接线时,efficientAD和本文方法的局部分支热力图都成功定位到了这个断裂的位置,但在全局分支中efficientAD错误地关注于连接头,并造成了最终的误检,RD的结果依然不如本文精确,而STPM没有定位到这个异常。在检测破碎的螺栓时,efficientAD和本文方法在局部分支中都成功定位了异常区域,在全局分支中,虽然两者都没有检测到这个结构异常,但本文方法相比于efficientAD产生的错误更少,减少了对结果的干扰,没有影响组合异常图的正确结果,RD产生了多余的响应,而STPM没有找到这个异常。在检测果汁瓶上的划痕缺陷时,本文方法相比于efficientAD在全局分支和组合异常图上检测区域更准确、误检更少,并且都优于STPM和RD的结果。以上结果证明了本文方法虽然是针对逻辑异常缺陷进行改进的,但对

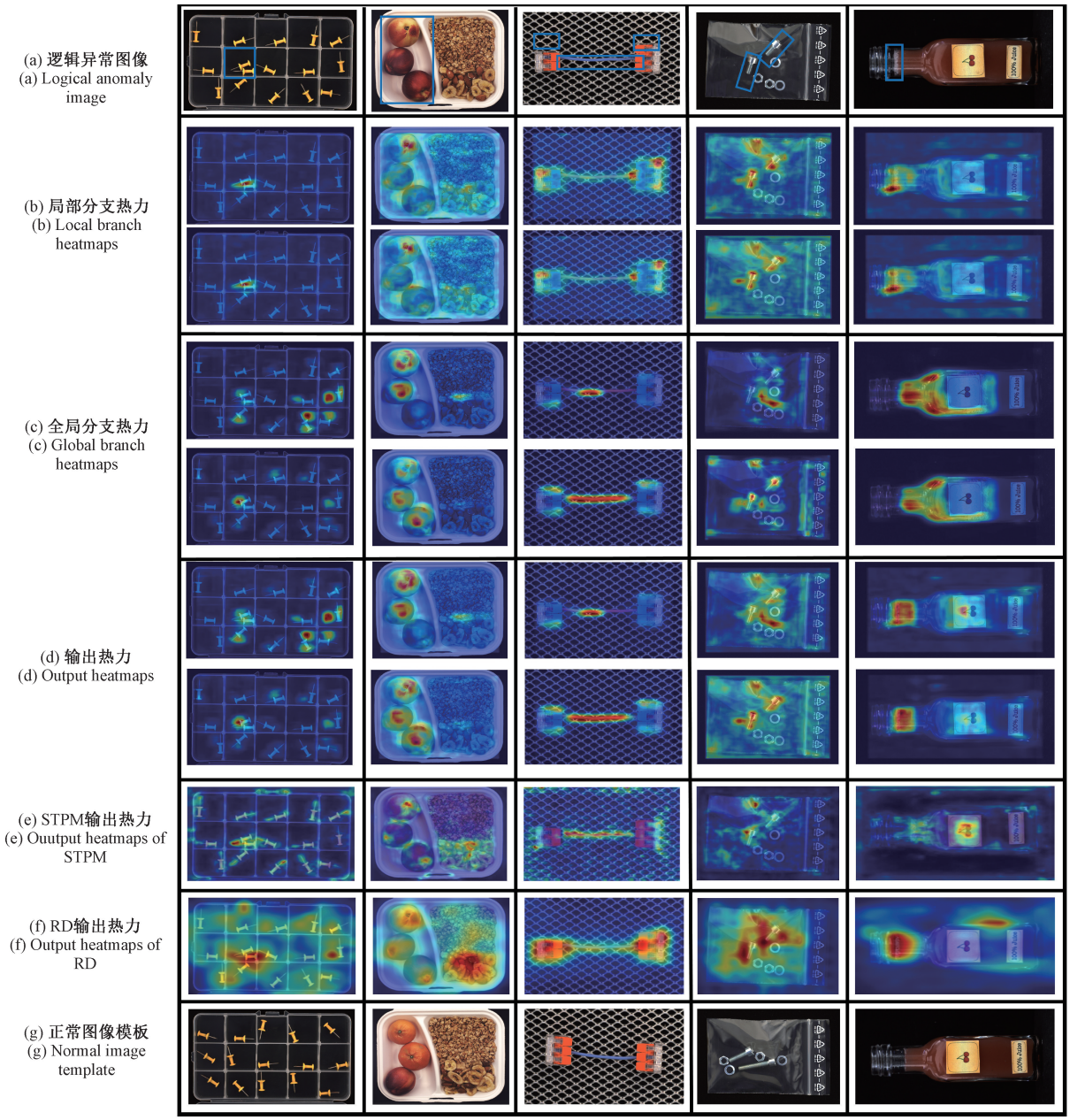


图 5　本文方法与 EfficientAD、STPM、RD 方法在逻辑异常图像上的可视化结果，在(b)、(c)、(d)分图中，EfficientAD 的结果位于上方，本文方法的结果位于下方

Fig. 5　The visual results of the proposed method, EfficientAD, STPM and RD on logical anomalous images, in sub-figures (b), (c) and (d), the results of EfficientAD are shown above and the result this research method are below

于检测结构异常仍然有帮助，这得益于 CSF 损失函数的设计，避免了局部分支网络检测性能的下降，也得益于并行双编码器架构改善了全局分支对结构异常缺陷的误检。

2.3　消融实验

为了验证本文方法中提出的各个模块和设计的有效性，本文进行了多组的消融实验进行对比。

1) SRM、CRM 和并行双 AE

表 2 对比了 SRM 模块和 CRM 模块和并行双 AE 结构的有效性。由实验 1、2 和实验 4、5 的对比可知，SRM 模块能有效地提高模型逻辑异常检测能力。在早餐盒类中，加入 SRM 的模型检测区域更加完整；在图钉类中，避免了很多区域的误检；并且在果汁类中，EfficientAD 模型没能检测出产品标签和果汁不匹配的逻辑异常，本文方法成功标出了标签的错误。类似由实验 1、3 和实验 4、6 对比，也可证明 CRM 模块的有效性，例如在连接管类中，本文方法检测到了更准确的区域，并同样能检测到果汁

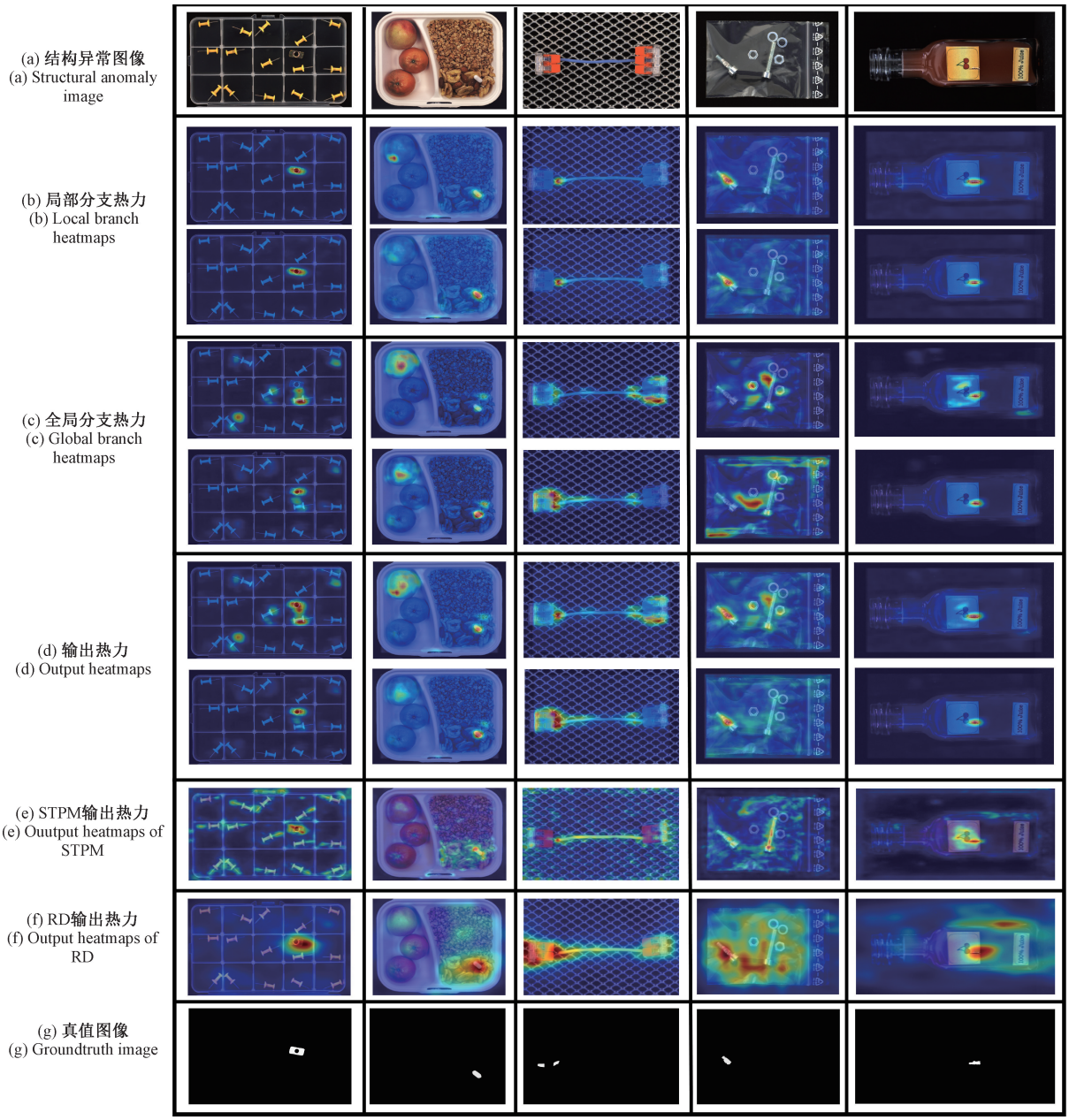


图 6　本文方法与 EfficientAD、STPM、RD 方法在结构异常图像上的可视化结果，在(b)、(c)、(d)分图中，EfficientAD 的结果位于上方，本文方法的结果位于下方

Fig. 6　The visual results of the proposed method, EfficientAD, STPM and RD on structural anomalous images, in sub-figures (b), (c) and (d), the results of EfficientAD are shown above and the results this research method are below

液体与标签的不匹配。SRM 和 CRM 模块的有效性得益于两者分别从空间通道两个角度提高了全局网络的感受野，学习了长程依赖关系。对比实验 1 和实验 4 结果可知，并行双 AE 结构能有效提高网络对逻辑异常的检测能力，并一定程度上提高了结构异常的检测能力。

对比实验 4、7 和图 7 可知，相比于基线方法在并行双 AE 结构下，SRM 模块和 CRM 模块结合使用能够进一步提高模型性能，例如在检测图钉盒中多出的图钉缺陷时，加入 CRM 模块的模型重点标注出了上方一处的图钉，加入了 SRM 模块的模型重点标注出了中间位置，并行双编码器标出了 3 处位置，这个结果是更合理的。在检测早餐盒中错误摆放的水果时，3 处水果的位置都有可能为缺陷区域，使用 CRM 和 SRM 模块的单编码器模型都只标注出了其中两处，而并行双编码器的检测结果

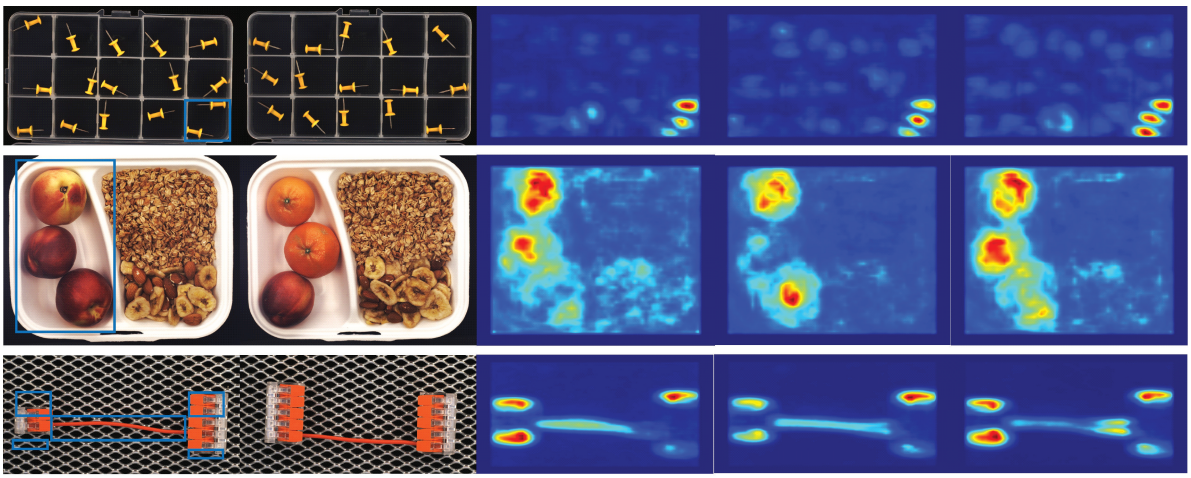
表 2　空间聚合模块、通道聚合模块和并行双 AE 结构的消融实验结果

Table 2　Ablation results of spatial reunion module, channel reunion module and parallel dual AE structure

实验	并行双 AE	SRM	CRM	平均(L/S)/%	平均/%
1(Baseline)				86.7/94.1	90.4
2		√		87.6/**94.2**	**90.9**
3			√	**87.8**/94.0	**90.9**
4	√			87.3/94.5	90.9
5	√	√		88.0/94.5	91.3
6	√		√	87.8/**94.8**	91.3
7	√	√	√	**88.2**/94.6	**91.4**

结合了两者的优势,检测更加完整。在检测连接器中连接线与接头数匹配错误的逻辑缺陷时,正常样本如果是红色连接线应对应两边都是 5 头的接头,这样应判断左侧的接头数错误,如果是黄色的连接线对应两边都是 2 头的接头,此时应判断连接线和右侧接头数错误,图 7 中可以看到加入 CRM 模块的检测结果对连接线和右侧接头检测不完整,而 SRM 模块的检测结果给了左侧的接头数较低的异常值,相比之下结合了空间通道聚合模块的网络检测结果最完整,证明了结合两个网络的并行双编码器结构能有效互补双方的不足,改善检测结果。

为了验证并行双编码器架构相较于串联单编码器的



(a)输入异常图像
(a)Input anomaly image

(b)正常图像模板
(b)Normal image template

(c)输出热力图(EfficientAD + CRM)
(c)Output heatmap (EfficientAD + CRM)

(d)输出热力图(EfficientAD + SRM)
(d)Output heatmap (EfficientAD + SRM)

(e)输出热力图(EfficientAD + SRM+CRM)
(e)Output heatmap (EfficientAD + SRM+CRM)

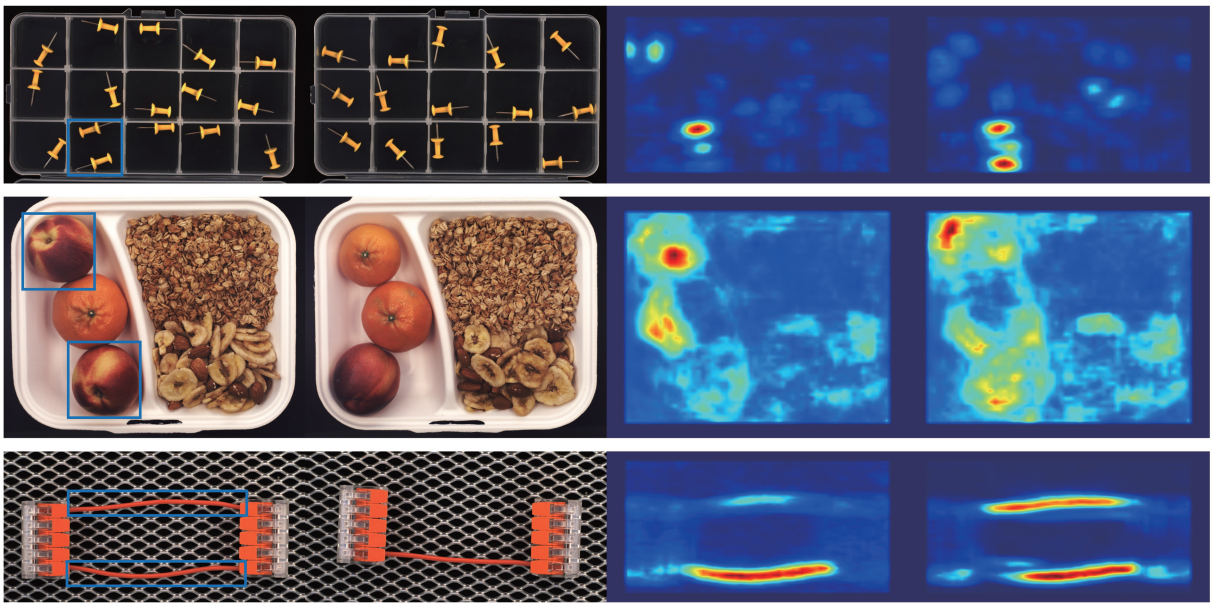
图 7　CRM 模块、SRM 模块在逻辑异常图像上的消融实验可视化结果

Fig. 7　The ablation visualization results of spatial reunion module and channel reunion module on logical anomaly images

优越性,本文在基线模型的基础上,在单编码器中串联了 CRM 模块和 SRM 模块,将实验结果与本文方法进行了可视化对比,如图 8 所示。从图 8 可以看出,在检测图钉盒中多出的图钉缺陷时,两处图钉的位置都有可能为缺陷,并行双编码器准确标出了两处位置,比单编码器考虑更全面,并且在正常区域产生的错误响应也比串联单编码器更少,误差更少;在检测早餐盒中错误的水果种类时,图示缺陷为将一个橘子错误地摆放为桃子,由于对摆放的位置没有严格要求,因此两个桃子的区域都有可能为异常区域,串联单编码器只标记出了一处区域,并在正确的橘子位置有了过高的响应,而使用并行双编码器避免了这个问题;在检测连接器中多出的连接线时,两根连接线都有可能是多余的 1 根,将两者都标记出是更加合理的,因此并行双编码器的检测结果比串联单编码器更完整。这表明采用并行双编码器的架构相比于串联单编码器在提取长程语义信息、获取图像全局理解方面更具优势,有利于逻辑异常的检测。

2) SFM 和 CSF 损失函数

表 3 验证了 SFM 模块和 CSF 损失函数的有效性。实验 9、11 分别在实验 7、10 的基础上加入了 SFM 模块,对比结果可知 SFM 模块能实现高效的特征融合,避免了长程语义信息损失,使模型在逻辑异常上的检测精度有了较大的提升。本文分别在基线方法(实验 1)和实验 7 的基础上修改损失函数,即对比实验 1、8,实验 7、10 和实验 9、11 可知,CSF 损失函数使模型在结构异常和逻辑异常上的检测能力都有所提升。同时,结合本文提出的模块与损失函数能实现互补作用,达到最好的检测效果。



(a)输入图像（逻辑异常）
(a)Input image (logical anomaly)

(b)正常图像模板
(b)Normal image template

(c)输出热力(串联单自编码器)
(c)Output heatmap (Series single autoencoder)

(d)输出热力(并联双自编码器)
(d)Output heatmap (Parallel dual autoencoders)

图 8　并行双编码器与串联单编码器模型在逻辑异常图像上的消融实验可视化结果

Fig. 8　The ablation visualization results of parallel dual AE and single AE model in series on logical anomaly images

表 3　SFM 模块和 CSF 损失函数的消融实验结果

Table 3　Ablation results of SFM and CSF loss function

实验	双 AE+SRM+CRM	SFM	CSF loss	平均(L/S)/%	平均/%
1(Baseline)				86.7/94.1	90.4
8			√	**87.0/94.6**	**90.8**
7	√			88.2/94.6	91.4
9	√	√		89.2/94.7	91.9
10	√		√	88.4/94.7	91.5
11	√	√	√	**89.4/94.9**	**92.1**

3　结　论

与结构异常仅要求模型关注局部特征差异不同,逻辑异常的检测要求模型能够理解图像长程的约束关系,获取全局语义信息对于逻辑异常的检测起关键作用。针对这一需求,本文首先设计了并行空间通道双自编码器架构,包含 SRM 模块和 CRM 模块,从空间和通道两个方向分别得到包含全局空间和通道理解的特征,丰富模型对局部特征的上下文建模,其次为了避免两个分支的输出融合时的信息损失,提出了一个 SFM 模块,实现了高效的信息交互,提高网络对逻辑异常的检测效果。最后在自编码器与学生网络的训练过程中设计了 CSF 损失函数,使网络更好地捕捉重要的上下文信息。在真实工业数据集上的实验结果,本文方法大幅提高了逻辑异常检测效果,并实现了结构异常和逻辑异常检测的兼顾,优于其他现有方法。

参考文献

[1] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks [J]. Medical Image Analysis, 2019, 54: 30-44.

[2] 袁磊, 唐海, 陈彦蓉, 等. SGCNet:一种轻量化的新能源汽车电池集流盘缺陷检测模型[J]. 电子测量与仪器学报,2023,37(10):172-182.
YUAN L, TANG H, CHEN Y R, et al. SGCNet: A lightweight defect detection model for new energy vehicle battery collector tray [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(10): 172-182.

[3] 吴一全, 赵朗月, 苑玉彬, 等. 基于机器视觉的 PCB 缺陷检测算法研究现状及展望[J]. 仪器仪表学报, 2022,43(8):1-17.
WU Y Q, ZHAO L Y, YUAN Y B, et al. Research status and the prospect of PCB defect detection algorithm based on machine vision [J]. Chinese Journal of Scientific Instrument,2022,43(8):1-17.

[4] RIPPEL O, MERTENS P, MERHOF D. Modeling the distribution of normal data in pre-trained deep features for anomaly detection [C]. 2020 25th International

Conference on Pattern Recognition (ICPR). IEEE, 2021: 6726-6733.

[5] RUFF L, VANDERMEULEN R, GOERNITZ N, et al. Deep one-class classification [C]. International Conference on Machine Learning, 2018: 4393-4402.

[6] BAUR C, WIESTLER B, ALBARQOUNI S, et al. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images [C]. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4. Springer International Publishing, 2019: 161-169.

[7] YANG H, CHEN Y F, SONG K Y, et al. Multiscale feature-clustering-based fully convolutional autoencoder for fast accurate visual inspection of texture surface defects[J]. IEEE Transactions on Automation Science and Engineering, 2019, 16(3): 1450-1467.

[8] COHEN N, HOSHEN Y. Sub-image anomaly detection with deep pyramid correspondences[J]. ArXiv preprint arXiv:2005.02357, 2020.

[9] DEFARD T, SETKOV A, LOESCH A, et al. Padim: a patch distribution modeling framework for anomaly detection and localization[C]. International Conference on Pattern Recognition. Cham: Springer International Publishing, 2021: 475-489.

[10] YU J W, ZHENG Y, WANG X, et al. Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows [J]. ArXiv preprint arXiv: 2111.07677, 2021.

[11] BERGMANN P, FAUSER M, SATTLEGGER D, et al. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 4183-4192.

[12] ZHANG X, LI SH Y, LI X, et al. DeSTSeg: Segmentation guided denoising student-teacher for anomaly detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 3914-3923.

[13] RUDOLPH M, WEHRBEIN T, ROSENHAHN B, et al. Asymmetric student-teacher networks for industrial anomaly detection[C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023: 2592-2602.

[14] WANG G D, HAN SH M, DING ER R, et al. Student-teacher feature pyramid matching for anomaly detection[J]. ArXiv preprint arXiv:2103.04257, 2021.

[15] BERGMANN P, BATZNER K, FAUSER M, et al. Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization [J]. International Journal of Computer Vision, 2022, 130(4): 947-969.

[16] GUO H W, REN L P, FU J J, et al. Template-guided hierarchical feature restoration for anomaly detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 6447-6458.

[17] DAI S M, WU Y F, LI X Q, et al. Generating and reweighting dense contrastive patterns for unsupervised anomaly detection [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(2): 1454-1462.

[18] KIM S, AN S, CHIKONTWE P, et al. Few shot part segmentation reveals compositional logic for industrial anomaly detection [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(8): 8591-8599.

[19] CHEN Z X, XIE X H, YANG L X, et al. Hard nominal example-aware template mutual matching for industrial anomaly detection [J]. ArXiv preprint arXiv: 2303.16191, 2023.

[20] YAO H, YU W, LUO W, et al. Learning global-local correspondence with semantic bottleneck for logical anomaly detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 34(5): 3589-3605.

[21] ZHANG J, SUGANUMA M, OKATANI T. Contextual affinity distillation for image anomaly detection [C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024: 149-158.

[22] DENG H, LI X. Anomaly detection via reverse distillation from one-class embedding[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 9737-9746.

[23] BATZNER K, HECKLER L, KÖNIG R. Efficientad: Accurate visual anomaly detection at millisecond-level latencies[C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024: 128-138.

[24] DOSOVITSKIY A, BROX T. Generating images with perceptual similarity metrics based on deep networks[J]. Advances in Neural Information Processing Systems, 2016, 29, DOI:10.48550/arXiv.1602.02644.

[25] WU L, WANG Y, GAO J B, et al. Deep adaptive feature embedding with local sample distributions for person re-identification[J]. Pattern Recognition, 2018, 73: 275-288.

[26] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. ArXiv preprint arXiv: 1412.6980, 2014.

[27] COHEN N, TZACHOR I, HOSHEN Y. Set features for fine-grained anomaly detection [J]. ArXiv preprint arXiv:2302.12245, 2023.

[28] YAO H, LUO W, YU W. Visual anomaly detection via dual-attention Transformer and discriminative flow[J]. ArXiv preprint arXiv:2303.17882, 2023.

作者简介

梁宵,2022 年于南京林业大学获得学士学位,现为江南大学硕士研究生,主要研究方向为深度学习、异常检测。

E-mail: 937126401@qq.com

Liang Xiao received his B. Sc. degree from Nanjing Forestry University in 2022. Now he is a M. Sc. candidate of Jiangnan University. His main research interests include deep learning and anomaly detection.

陈莹(通信作者),2005 年于西安交通大学获得博士学位,现为江南大学教授、博士生导师,主要研究方向为机器视觉、信息融合、模式识别。

E-mail: chenying@jiangnan.edu.cn

Chen Ying (Corresponding author) received her Ph. D. degree from Xi′an Jiaotong University in 2005. Now she is a professor and Ph. D. supervisor in Jiangnan University. Her main research interests include machine vision, information fusion and pattern recognition.